Age-related differences in richness and diversity of Russian color lexicon

Yulia A. Griber [1,*], Dimitris Mylonas [2] and Galina V. Paramei [3]

[1] Smolensk State University, Smolensk, Russia; y.griber@gmail.com
[2] New College of the Humanities at Northeastern University, London, UK; dimitris.mylonas@nchlondon.ac.uk
[3] Liverpool Hope University, Liverpool, UK; parameg@hope.ac.uk
***** Corresponding author: y.griber@gmail.com

Abstract

In the present study, we investigated age-related differences in richness and diversity of color lexicon in Russian native speakers. Color names were elicited in 2018–2020 in an ongoing web-based psycholinguistic experiment (https://colournaming.com). An unconstrained color-naming method was employed. A final dataset contained responses of 1,967 native Russian speakers (1,280 females, 677 males, 10 non-binary), from various locations of the Russian Federation, aged 16–98 years. Participants were stratified into seven age groups (years): 16–19 (M=17.59 ± 1.19), 20–29 (M=24.86 ± 3.11), 30–39 (M=35.04 ± 2.89), 40–49 (M=45.18 ± 2.86), 50–59 (M=55.63 ± 2.90), 60–69 (M=65.05 ± 2.87) and ≥70 (M=78.20 ± 4.97). To estimate heterogeneity of color lexicon in each age group, we applied the Margalef and Simpson indices broadly used for measuring ecological diversity. The indices enabled gauging richness of color lexicon, i.e., the number of word types in the dataset, and color-term evenness, i.e., the relative abundance of different color terms. Our analysis of synchronic variability provides evidence that color vocabulary develops actively throughout the entire adult life and remains relatively stable in both richness and diversity up to the old age. Respondents of the three younger groups, under 40 years, revealed the greatest color lexicon diversity. In comparison, in the 40–59-year-old the diversity index was lower, and decreased dramatically in respondents of 60 years and older. The apprehended dynamics reflects intergenerational differences as such, but even more so dramatic changes of sociocultural reality in the post-Soviet era (after 1991).

Keywords: *color naming, Russian, age-related differences, richness of lexicon, diversity of lexicon*

INTRODUCTION

Aging is known to affect various levels of language processing including but not limited to mental lexicon (Wulff et al. 2019), language variation (Pichler et al. 2018), lexical diversity and vocabulary size (Brysbaert et al. 2016), object naming ability and lexical retrieval (Connor et al. 2004).

Yet, there is surprisingly little research that explored variation of color lexicon in speakers of different age groups in a certain language. Among rare exceptions is a study of Zaręba (1954), who explored intergenerational differences of color names in Polish dialects. For non-industrialized cultures, Kay (1975) and Dougherty (1977) found convincing evidence that younger speakers use more color terms and manifest linguistically more refined color space partitioning than older speakers. More recently, in older Udmurt speakers Ryabina (2009) recorded many more elaborate color terms – qualified basic color terms (BCTs), non-BCTs or qualified fancy terms. For Swedish speakers, Vejdemo (2018) revealed that older Swedes used many more modifiers and color compounds than younger speakers; moreover, there was an intergenerational flux in labeling basic color categories, with lexical replacement demonstrated by younger speakers.

In the present study, we investigated age-related differences in richness and diversity of color lexicon in Russian native speakers.

MATERIALS AND METHODS

Web-based psycholinguistic experiment

Color names were elicited in 2018–2020 in an ongoing web-based psycholinguistic experiment (https://colournaming.com). Color stimuli (N=606) were approximately uniformly distributed in the Munsell color space. An unconstrained color-naming method was employed: observers were presented with randomly ordered stimuli, one at a time, and asked to name the color by typing their responses in Russian using Cyrillic alphabet. Respondents were free to use any color descriptor – either a single word, or a compound, or term(s) with modifiers or qualifiers (for further details see Mylonas and MacDonald 2010; Griber et al. 2018; Paramei et al. 2018).

Participants

A final dataset contained responses of 1,967 native Russian speakers (1,280 females, 677 males, 10 non-binary), from various locations of the Russian Federation, aged 16–98 years (Table 1).

	Total	**16–19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70+**
Female	1,280	74	324	253	179	165	165	120
Male	677	19	127	139	101	88	121	82
Other	10	2	2	1	2	1	2	0
Total	1,967	95	453	393	282	254	288	202

Table 1: Number of participants of different genders, in the total sample and stratified into seven age groups.

Participants were stratified into seven age groups (Table 2).

	Total	**16–19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70+**
Mean age	42.54	17.59	24.86	35.04	45.18	55.63	65.05	78.20
SD	17.71	1.19	3.11	2.89	2.86	2.90	2.87	4.97

Table 2: Mean (SD) age of participants, in the total sample and stratified into seven age groups.

The refined dataset contained 55,516 responses that undergone linguistic analysis (Table 3).

	Total	**16–19**	**20-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70+**
Number of responses	55,516	2,175	11,581	10,287	7,107	7,297	6,192	3,864
Number of word types	3,128	354	1299	1034	840	740	385	209

Table 3: Number of unique responses and word types.

Diversity indices

To estimate heterogeneity of color lexicon in each age group, we applied the indices broadly used in ecological studies (where diversity of species is crucial for evaluating health of ecological systems).

The diversity indices can be of two types: those that assess species' (1) richness (the number of species), and (2) evenness, or dominance (the distributive number of individual organisms among the species).

Analyzing individual color names by analogy with biological species, we used diversity measures of both types, to account for color lexicon richness (the number of color terms, or word types, present in the dataset), as well as color term evenness (relative abundance of the different color terms) for each age group.

To assess color lexicon richness, we applied the Margalef (1958) diversity index that captures system's entropy. Here, to calculate the diversity index for each age group, we used the formula suggested by Iglesias-Rios and Mazzoni (2014):

$$R = (s–1) / \ln N, \qquad (1)$$

where *s* is the number of word types, and *N* is the total number of responses in the sample.

The Margalef index enables to capture the number of different color names (word types) in the dataset of the participant sample. The more word types are present, the 'richer' is the sample's color lexicon. Note that the Margalef index does not take into account the number of responses in each word type: it allocates equal weight both to those color names that occurred very rarely and those that occurred frequently among participants' responses. For instance, color name *belosnežnyj* 'snow-white' offered by a single participant has as much influence on the richness measure as BCT *fioletovyj* 'purple' offered thousands of times.

To assess evenness, i.e., relative abundance of different color names in the dataset, we implemented the Simpson index (Simpson, 1949) defined as:

$$D = 1 – \Sigma(n_i (n_i–1)) / N(N–1), \qquad (2)$$

where n_i is the number of responses of the i_{th} word type, and *N* is the total number of responses in the dataset.

Thus, the Simpson index takes into account not only the number of word types present in the dataset, but also the number of occurrences of each word type. Its value ranges between 0 and 1, and presents the probability that two responses randomly selected from the dataset will contain different types of color names.

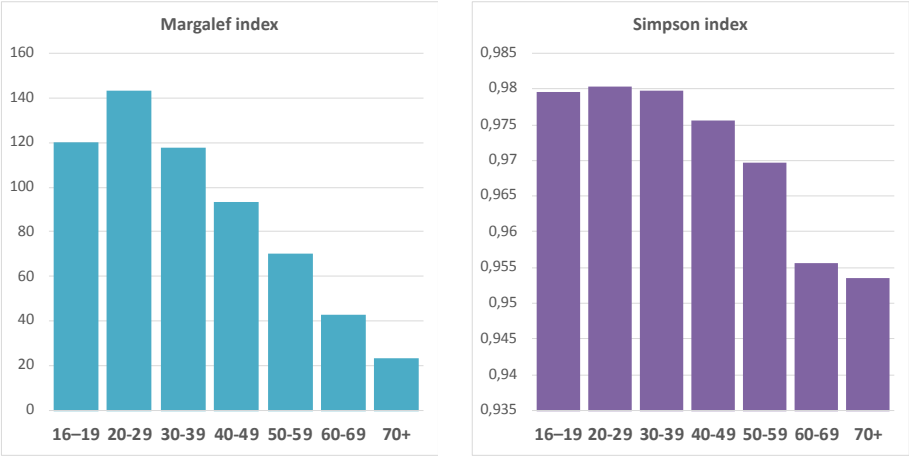

Figure 1: The Margalef (left) and Simpson (right) diversity indices for the seven age groups.

RESULTS

Our apparent-time analysis (or synchronic variability), gauged by the two indices of diversity, provides evidence that color vocabulary develops actively throughout the entire adult life and remains relatively stable in both richness and diversity up to the old age (Figure 1).

Since the Margalef index is highly sensitive to the sample size (here: size of the age group), we calculated it, in addition, for each year-of-life cohort, to measure color lexicon diversity of the 16-year-old, 17-year-old and so on (Figure 2).

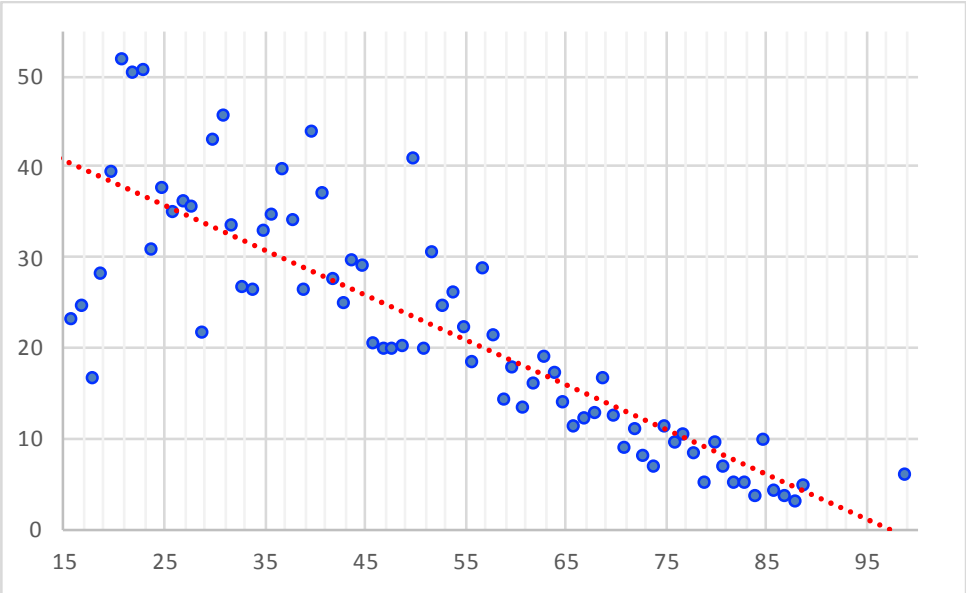

Figure 2: The Margalef diversity index estimated for each year-of-life cohort in the sample aged 16–98 years old.

The Simpson index (Figure 1, right) indicates that respondents of the three younger groups, under 40 years, revealed the greatest color lexicon evenness. Furthermore, color inventories of the 16–19-year-old and of those in their 20s and 30s are rather different from color inventories of other age groups. In particular, along with 12 Russian BCTs, the younger groups' color vocabulary was richer and more variegated, and included abundant monolexemic non-BCTs, modified and compounded color terms.

In comparison, in the 40–59-year-old the Margalef richness index was lower, and decreased dramatically in respondents of 60 years and older (Figure 1, left, Figure 2). The decline in both richness and evenness of color lexicon observed here in mature speakers is in accord with the findings of general waning, with age, of the vocabulary size and lexical diversity (see, e.g., Brysbaert et al. 2016).

Beyond the diversity measures, we observe further ongoing juniority effects of color lexicon enrichment. For younger Russian speakers, these effects have three main manifestations:

(i) Augmenting of the inventory of basic color terms;

(ii) Active lexical refinement of four basic color categories, PURPLE, GREEN, BLUE and PINK, and, as well, of "hard-to-name" areas of color space at the boundaries of BLUE-GREEN, YELLOW-GREEN and ORANGE-BROWN categories;

(iii) Qualitative differences in younger generations' lexicon, specifically: accruing of novel terms whose object references emerged in the last three decades; an increasing use of idiosyncratic terms that serve an expressive rather than informative function.

CONCLUSIONS

Our results provide evidence of intergenerational differences in color lexicon in the representative sample of native Russian speakers characterized by a significant age span, from 16 to 98 years old, and stratified into seven age groups.

The Margalef and Simpson diversity indices were applied for analyzing color term inventories for individual age groups. In tandem, the diversity indices enabled gauging and comparing richness of color lexicon (i.e., the number of word types) and color term evenness (i.e., relative abundance of different color terms) in each of the seven age groups.

We found both quantitative and qualitative differences in the color word types between younger generations (aged 16–39 years) and those aged 40 and older: color inventory of the former is both richer and more diverse. The apprehended apparent-time dynamics, namely, the juniority effects of color-term incrementation, reflects intergenerational differences as such, but even more so dramatic changes of sociocultural reality in the post-Soviet era (after 1991), whereby Russian speakers became greatly impacted by globalization of trade with new market product arrivals. This resulted in elaboration of novel (frequently loan) terms, especially by younger speakers, for efficient communication about perceived color of the chromatic environment significantly enriched by coloration of new diverse products.

ACKNOWLEDGEMENTS

Y.G. was supported by the Grant 17-29-09145 of the Russian Foundation for Basic Research. D.M. was supported by a FY22 TIER 1 Seed Grant from Northeastern University (US). Support of Tatjana Samoilova, Alexei Delov, and Karina Zygankova in data processing is gratefully appreciated. The authors are also grateful to all participants for their time and good will.

REFERENCES

Brysbaert, M., M. Stevens, P. Mandera, and E. Keuleers. 2016. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology 7*: 1116. doi: 10.3389/fpsyg.2016.01116

Connor, L. T., A. Spiro, L. K. Obler, and M. L. Albert. 2004. Change in object naming ability during adulthood. *The Journals of Gerontology, Series B 59*(5): 203-209.

Dougherty, J. W. D. 1977. Color categorization in West Futunese: Variability and change. In: *Sociocultural dimensions of language change*, B. G. Blount, ed. New York: Academic Press, 143-148.

Griber, Y. A., D. Mylonas, and G. V. Paramei. 2018. Objects as culture-specific referents of color terms in Russian. *Color Research and Application 43*(6): 958-975.

Iglesias-Rios, R., and R. Mazzoni. 2014. Measuring diversity: Looking for processes that generate diversity. *Natureza & Conservação* 12(2): 156-161.

Kay, P. 1975. Synchronic variability and diachronic change in basic color terms. *Language in Society 4*(3): 257-270.

Margalef, R. 1956. Información y diversidad específica en las comunidades de organismos [Information and specific diversity in communities of organisms]. *Investigación Pesquera* 3: 99-106 (in Spanish).

Mylonas, D., and L. MacDonald. 2010. Online colour naming experiment using Munsell samples. *Proceedings of the 5th European Conference on Colour in Graphics, Imaging, and Vision – CGIV 2010,* 27-32.

Paramei, G. V., Y. A. Griber, and D. Mylonas. 2018. An online color naming experiment in Russian using Munsell color samples. *Color Research and Application 43*(3): 358-374.

Pichler, H., S. E. Wagner, and A. Hesson, A. 2018. Old-age language variation and change: Confronting variationist ageism. *Language and Linguistics Compass 12*(6): 1-21.

Ryabina, E. 2009. Sex-related differences in the colour vocabulary of Udmurts. *Wiener elektronische Beitrage des Instituts für Finno-Ugristik (WEBFU)*. http://webfu.univie.ac.at/texte/12Ryabina.pdf

Simpson, E. H. 1949. Measurement of diversity. *Nature 163*: 688. https://doi.org/10.1038/163688a0

Vejdemo, S. 2018. Lexical change often begins and ends in semantic peripheries: Evidence from color linguistics. *Pragmatics & Cognition 25*(1): 50-85.

Wulf, D. U., S. De Deyne, M. N. Jones, and R. Mata. 2019. The Aging Lexicon Consortium. New perspectives on the aging lexicon. *Trends in Cognitive Sciences 23*(8): 686-698.

Zaręba, A. 1954. *Nazwy barw w dialektach i historii języka polskiego [Color names in dialects and history of the Polish language]*. Wrocław: Ossolińskich (in Polish).